J Clin Epidemiol Vol. 43, No. 3, pp. 285–295, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

A LARGE-SCALE PROSPECTIVE COHORT STUDY ON DIET AND CANCER IN THE NETHERLANDS

PIET A. VAN DEN BRANDT,[1]* R. ALEXANDRA GOLDBOHM,[2] PIETER VAN 'T VEER,[2] ALEXANDER VOLOVICS,[3] RUDOLPH J. J. HERMUS[2] and FERD STURMANS[1]

[1]Department of Epidemiology, University of Limburg, Maastricht, The Netherlands, [2]Department of Nutrition, TNO-CIVO Toxicology and Nutrition Institute, Zeist, The Netherlands and [3]Department of Medical Statistics, University of Limburg, Maastricht, The Netherlands

(*Received in revised form* 29 *June* 1989)

Abstract—In 1986, a prospective cohort study on diet and cancer was started in The Netherlands. The cohort ($n = 120, 852$) of 55–69 year old men (48.2%) and women (51.8%) originates from 204 computerized municipal population registries. At baseline, participants completed a self-administered questionnaire on diet and potential confounding variables. In addition, about 67% of the participants provided toenail clippings. Cancer follow-up consists of record linkage to a pathology registry and to cancer registries. The initial interest is in stomach, colorectal, breast and lung tumors. A case–cohort approach is applied, in which detailed follow-up information of a random subcohort ($n = 5000$) provides an estimate of the person-time experience of the cohort. Exposure data of the subcohort will be combined with those of incident cases, yielding exposure-specific incidence rate ratios. The intraindividual variation in determinants is estimated by annually repeated measurements ($n = 250$) within the subcohort. The rationale, efficiency aspects and study characteristics are discussed.

Diet Neoplasms Epidemiologic methods Biometry Questionnaires Toenails

INTRODUCTION

The possible role of dietary factors in the etiology of human cancer continues to be a subject both of research and debate. Various estimates have been produced on the proportion of cancer cases attributable to diet and other factors [1–3]. It has been rather difficult, however, to identify specific elements of the diet as being causative or preventive. Analytical epidemiological studies on diet and cancer have been mostly of the case–control type; their results often seem to lack consistency, which may be attributed partly to the potential for selection bias, and, particularly in dietary studies, recall bias [4]. Considering that large-scale randomized controlled dietary intervention trials are rarely feasible (because of financial, blinding, compliance and ethical reasons), prospective cohort studies are often proposed as the alternative method of choice. At the same time cohort studies are commonly regarded as prohibitively expensive, notably studies among the general population. The costs generally originate from recruitment of the study population, (baseline) exposure measurement and follow-up. Thus, there is a need for cost-efficient prospective cohort studies [4].

Various ongoing cohort studies on diet and cancer have been published, with widely differing characteristics. The following serves merely as a general description of the characteristics, supplemented with some examples of studies, without attempting to be complete.

*Reprint requests should be addressed to: Piet A. van den Brandt, Department of Epidemiology, University of Limburg, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

Most studies are conducted in the general population (e.g. [5, 6]) or in a captive (occupational) subgroup of specific gender (e.g. [7]); other studies use groups with specific dietary habits (e.g. [8–10]). Studies may have been started as such or have been attached to existing data collection structures as a census (e.g. [11]) or to a screening program (e.g. [12, 13]); others have made use of existing biological banks, usually without any dietary assessment (e.g. [14, 15]. Dietary assessment may include a limited (e.g. [11, 16]) or more extensive questionnaire [7, 17], with the length usually inversely related to sample size. Still other studies have used a single 24 hr dietary recall [18, 19] or cross-check dietary history interviews [20]. Because of their characteristics, some studies will yield data on overall dietary habits, which can hardly be translated into nutrient intakes. In addition, it might be difficult to generalize such results from studies when conducted in Japan, to Western Europe or the U.S. Studies that address the issue of nutrient intake or use biochemical markers generally take a relatively long time to yield a large number of cases because of the small study size. Indeed, a combination of dietary assessment and biological sampling on a large scale is rare [7], whereas both sources of information are likely to be complementary.

Our objective was to design an efficient, large-scale study among men and women that combines extensive dietary assessment with biological sampling, and that yields a sufficiently large number of cases within a reasonably short follow-up period. Efficiency in this respect refers to selection of the study population and area, determinant contrasts and their measurement, biological sampling, follow-up, data processing and statistical analyses. The cohort study was started in 1986 and was preceded by a pilot study in 1984 and 1985 to evaluate the feasibility of the project and develop the methods to be used. After presenting the general outline of the study [21], the various design considerations and decisions will be discussed in detail.

GENERAL OUTLINE OF THE COHORT STUDY

The primary purpose of the study is to investigate the effects of fats, vitamins, fiber, alcohol, selenium, nitrate, sodium and calcium on the development of gastric, colorectal, breast and lung tumors. Cancer risk associated with specific dietary patterns will also be evaluated. These tumor sites chosen because of their suggested relationship with dietary factors (e.g. [22, 23]) and their high incidence in The Netherlands [24].

The study is conducted among 55–69 year old men and women. Subjects originate from the general population sampled from municipal population registries. The pilot study indicated that a fairly large contrast in dietary intake exists in this population. To increase the contrast in the cohort still further, individuals with special dietary habits (e.g. vegetarians) are overrepresented. Information on determinants is obtained by a self-administered questionnaire and collection of toenail clippings. The 11-page questionnaire contains 6 pages on food habits, supplemented with questions on potential confounders and other independent risk factors. These include: smoking and occupational history, socioeconomic status, history of selected medical conditions, family history of cancer, chronic drug use, reproductive history, obesity and physical activity. The cohort is constituted by the 120,852 subjects who completed the baseline questionnaire that was sent to a total of 340,439 subjects.

Follow-up for cancer incidence will be performed by record linkage to PALGA (a data base on Dutch pathology reports) and to the cancer registries. During the first 5 years of follow-up, approximately 250 cases of stomach cancer, 450 colon, 300 rectal, 800 breast and 1200 lung cancer cases are expected to arise from this cohort, taking mortality into account [25, 26].

A case–cohort approach is applied, by selecting a random subcohort ($n = 5000$) from the large cohort immediately after identification of the cohort members. This subcohort is being followed up for migration and vital status by contacting the participants and the municipalities. As will be discussed, for testing the primary study hypotheses a subcohort size of 3500 is sufficiently large. Therefore, questionnaires and toenail specimens are initially processed only for a random subsample of 3500 out of the 5000 subcohort members. However, the person-time experience is also collected for the remaining 1500 subjects, whose covariate data will be processed when hypotheses regarding rare exposures responsible for a small proportion of specific tumors are of interest [27]. In the statistical analyses using the proportional hazards model [28], stratification on year of follow-up will be employed to investigate the

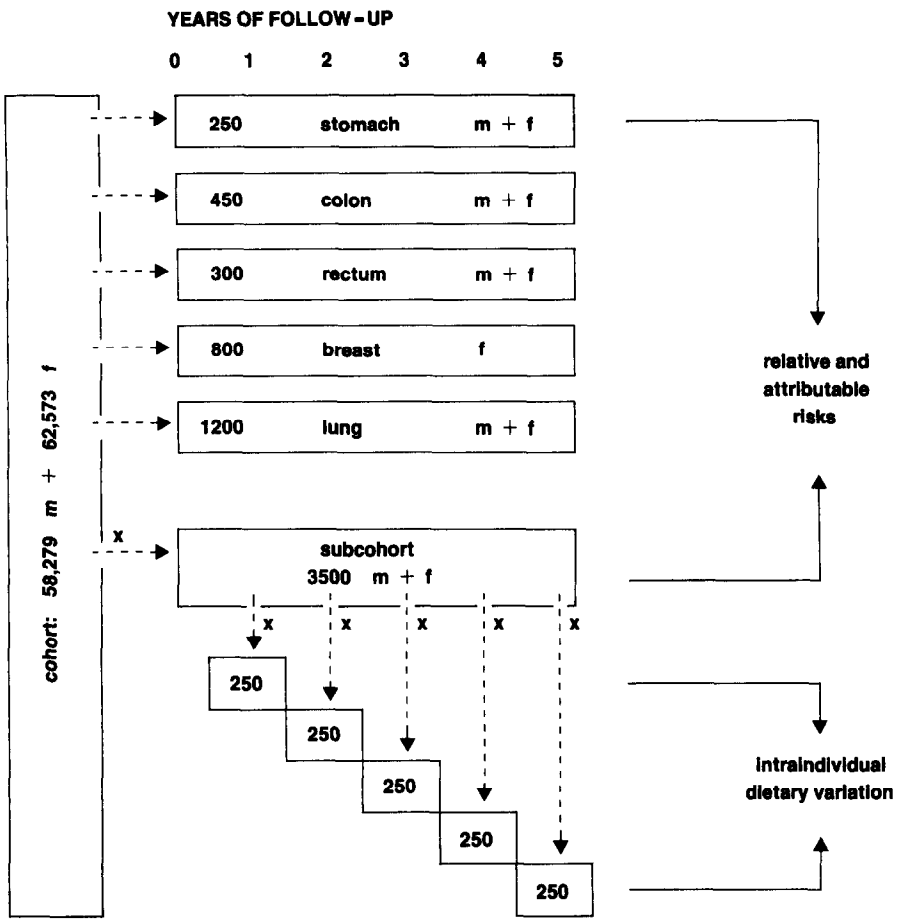

Fig. 1. Design aspects of the prospective cohort study. x, random sample; m, males; f, females.

influence of possible preclinical disease at the start of the study. The intraindividual variation in determinants will be estimated by repeating the questionnaire annually in subsamples ($n = 250$ each) of the subcohort. These design aspects are depicted in Fig. 1.

SELECTION OF THE STUDY POPULATION

Age at entry

Because it has been suggested that diet predominantly exerts its role on the later stages rather than the early stages of tumorigenesis [23, 29], this cohort study is conducted among elderly people aged 55–69 years at entry. Also, younger individuals generally show less stable dietary habits, because they tend to consume more new foodstuffs [30]. In two longitudinal Dutch studies, changes in nutrient intake over a 3 year period have been studied using interperiod correlation coefficients. The decrease in correlation among women around menopause [31] was much smaller than among adolescents [32]. Although adolescents are an extreme category in this respect, this comparison provides some evidence for higher stability at older ages.

In the age group well above 70 years, problems may occur with the dietary assessment, and there is a tendency for underreporting and less histological verification of elderly cancer patients. Finally, a relatively short follow-up (5 years) of a large cohort from the selected age stratum will yield a sufficient number of cases to perform meaningful statistical analyses, i.e. minimally 300 cases per tumor site.

Size and area

In order to obtain this number of cancer cases, it was estimated that a cohort size of 150,000 subjects is required [33]. The choice of the study population is then determined largely by recruitment efficiency and the required accuracy of identifying information in view of the proposed method of follow-up (i.e. record linkage). In The Netherlands, (computerized) municipal population registries contain highly accurate identifying information on every

citizen, and constitute an efficient sampling frame for the general population. Since a mailed data collection procedure would be used (with a lengthy questionnaire), the aim was to start with an initial sample size of 350,000 subjects in order to establish a cohort of about 150,000 respondents.

The study area was defined in terms of municipalities satisfying the following eligibility criteria: (a) availability of a computerized population registry; (b) sufficient cancer follow-up coverage.

In 1985, 323 out of the 714 municipalities were computerized; 300 (93%) of them agreed to provide in 1986 a gender-stratified random sample of specified size, equivalent to 40% of each municipal 55–69 years age stratum. Cancer follow-up coverage was determined as follows. Recently, two sources of incident cancer cases have become available: PALGA and 9 cancer registries. Since both PALGA and the cancer registries were not yet operating in the entire country, a list of collaborating hospitals (in 1986) was obtained. Together with data on the municipal origin of all patients admitted for cancer to these hospitals (obtained from the National Health Care Information Center), expected municipal follow-up coverage degrees were calculated per tumor site of interest. From the list of 300 computerized municipalities, 204 were selected with a coverage degree exceeding 75%, yielding a tentative initial sample of almost 340,000 people. The estimated mean coverage degree for cases of any of the 5 tumor sites of initial interest was 93% in this case. Loss to follow-up due to migration out of the coverage area (estimated at 1.9% in 5 years) is taken into account in this estimate. The location of the 204 selected municipalities is displayed in Fig. 2. Municipal samples were selected in May–August 1986, accumulating to 339,733 subjects.

Recruitment of subjects with special dietary habits

Apart from sampling and follow-up considerations, the expected exposure contrast (and

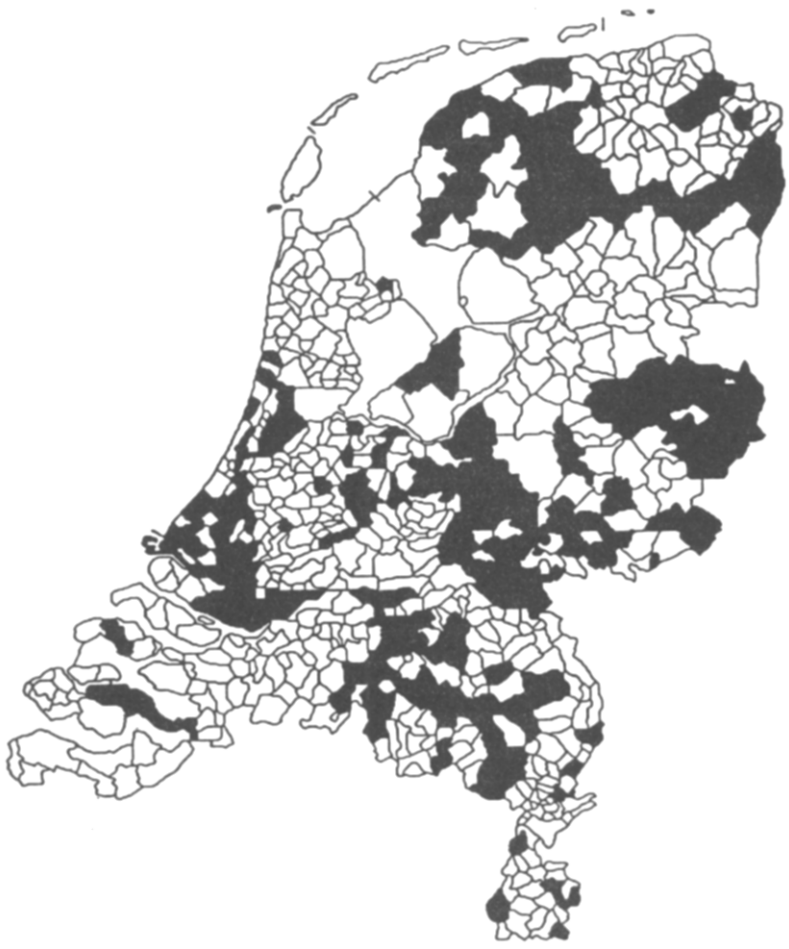

Fig. 2. The location of the participating municipalities in The Netherlands.

its temporal stability) in the selected study population needs to be contemplated. It has been suggested that in Western populations the contrast in exposure may be small when studying individuals within one region or country [34–36]. In our pilot study we have evaluated the dietary intake in the proposed study population, which showed a rather large dietary contrast (e.g. mean $\pm$ SD of dietary fat as percent of calories was $40.3 \pm 5.5\%$). Within western societies, this variation can be regarded as sizeable. It can be further augmented by extending the study to other countries with different dietary habits or by overrepresentation of individuals with deviant food habits. With the former approach, problems may arise with confounders and standardization of dietary assessment, in view of different eating patterns. These problems are less serious when the second approach is used. Therefore, subjects of 55–69 years and eating meat less than twice a week were invited to participate by advertisements and leaflets in life-style magazines and health food stores located in areas covered by PALGA or the cancer registries. During the period of recruitment (April–July 1986), about 1000 persons applied; 30% of all applicants were living outside the covered areas or did not have the correct age at entry. For some of the remaining subjects, extra contacts were needed to obtain complete and correct identifying information, even though standard application forms were used. Altogether, 706 eligible "vegetarians" were recruited in this way.

ASSESSMENT OF DIETARY EXPOSURE

Choice of dietary assessment method

Whereas a large interindividual variation in exposure is desirable and variation within subjects should preferably be minimal for observational etiologic studies, it is the ratio of intra- to-interindividual variation which determines whether meaningful contrasts in exposure can be studied. When this ratio is large, substantial random misclassification will result in attenuated measures of association between exposure and disease [37]. The variance ratio is in turn determined by the exposure characteristics of the study population, combined with the dietary assessment technique that is being used. In a transitional study population with rapidly changing dietary patterns as in Japan [38] and consequently a large interindividual variation in food intake, compared to intraindividual variation, a relatively simple method may suffice [11]. Compared to the rapid and dramatic changes in Japan, changes in per capita food intake in The Netherlands are moderate [39]. Our study population will therefore show less heterogeneity (but also relatively stable dietary habits). In order to minimize misclassification, this requires an elaborate dietary assessment method with a reference period of one year, covering seasonal variations. The dietary history can be used for this purpose, but the interview method is laborious and impractical in large-scale studies. Therefore, abbreviated methods like the (semiquantitative) food frequency questionnaire (SFFQ) have been developed, which can be self-administered. The validity and reproducibility of the SFFQ have been studied recently [40–43]. Combined with its feasibility, these results make it the method of choice for this cohort study. Furthermore, by repeating applications of the method annually to samples of the cohort, estimates can be made on the intraindividual variance in annual intakes. These estimates might also be used to improve estimates of the rate ratios and associated confidence intervals [44, 45].

Construction of the dietary questionnaire

A prerequisite for the development was that the questionnaire should be aimed at measuring the contrasts in dietary intake that exist in the cohort and it should be self-administrable. The construction of the questionnaire is described in more detail by Bausch-Goldbohm *et al.* [46]. Briefly, in 1984 and 1985 detailed dietary history interviews (covering the preceding year) were conducted by trained dieticians in a group of 169 subjects (including 20 vegetarians) of similar age and gender structure as the cohort population. After calculation of the intakes of 15 nutrients of interest (related to the hypotheses), multiple regression analyses were employed together with residual analysis, to select those food items that predicted most of the interindividual variation in the nutrient intakes of interest, as measured by the dietary history. Furthermore, the need for including questions on portion sizes was also evaluated by this method. Finally, the remaining list was supplemented with some items in order to maintain a logical (dietary) structure in the questionnaire. The result was a 6-page dietary questionnaire of 175 food items, that explained the variance in nutrient intake as measured by the dietary history, ranging from 86% for vegetable fiber to

100% for alcohol. The validity of the final version, that was used in the cohort, is further being tested against the dietary record method and the dietary history method in ongoing substudies.

Choice of biochemical markers

Because of the potential problems associated with the assessment of food intake, the use of biochemical markers of dietary exposures has been proposed as an objective, "hard-evidence" alternative. Although the use of biological specimens like plasma seems attractive in that the biochemical markers address the nutritional status more precisely, they nevertheless suffer from some inherent problems as well. The marker may not properly reflect long term nutritional status (e.g. [47]); large intra-individual variations in the marker content may result in a high ratio of intra-to-interindividual variation (e.g. serum cholesterol or urinary sodium). For retrospective etiologic studies, various markers may be of less value since the tumor may have altered the marker level, as has been shown for plasma Se, vitamin E and retinol [48–50]. In prospective studies, the collection, storage and analyses of specimens may be prohibitively expensive, leading to smaller cohorts with decreased power and an increased risk of chance findings. A promising exception to this is toenail specimens. These reflect long term intake of several micronutrients (e.g. selenium or zinc [51, 52]), and the specimens can easily be collected, transported by mail and stored at room temperature [51]. Given these characteristics and the study size, we included the collection of toenail clippings in our study.

BASELINE EXPOSURE MEASUREMENT

Conduct of baseline measurement and response

In September 1986, the 340,439 selected subjects were invited by mail to complete the questionnaire and collect toenail clippings. To return their completed questionnaire, respondents were offered the choice of using a business reply number (used by 33% of respondents) or (preferably) to provide their own stamp (used by 67%). The acceptability of this approach had been tested in the pilot [53]. Several large municipalities had explicitly stated, for reasons of privacy protection, that the selected subjects could only be approached once, without use of reminders. To elevate the response rate, a nationwide publicity campaign accompanied the baseline survey. Completed questionnaires were returned by 120,852 subjects (response rate 35.5%; men 34.5, women 36.6%). An estimated 67% of the respondents also provided toenail specimens. The first page of the questionnaire was optically scanned to define the cohort, to check specific identifying information needed for future linkage (e.g. date of birth, twinship). This page also contained questions on the presence of cancer and other conditions, overall smoking habits and special food habits (i.e. vegetarianism, veganism, etc.).

Some baseline characteristics of the cohort

The cohort is composed of 58,279 men (48.2%) and 62,573 women (51.8%). To examine whether the response in our study had affected the determinant distributions (e.g. did primarily non- or ex-smokers respond?), an analysis of response rates was carried out as far as the available sample data on nonrespondents permitted. Also, data from the first page of the questionnaire were used. Table 1 shows the response rate according to age and degree of urbanization of municipality of residence. Table 2 shows the distribution of marital status, smoking habits and overall frequency of meat consumption in the total cohort. No data on these variables are available for the non-respondents, but for the first two variables national large-scale survey data do exist [54, 55].

Furthermore, after the cohort was identified, a random sample was selected in 1987 to validate the dietary questionnaire against the dietary record method, using 9 recording days evenly distributed over the year 1987/1988. Available data at this moment permit a comparison of the intake of several nutrients of cohort members with data from a recent national survey in which a 2-day dietary record was used [30]. Results for caloric intake and calorie providing nutrients are presented in

Table 1. Response rate to baseline measurement among men and women according to age and degree of urbanization

	Response rate (%)	
Variable	Men (n = 58,279)	Women (n = 62,573)
Age (yr)		
55–69	34.6	38.8
60–64	35.1	36.8
65–69	33.6	34.1
Urbanization of municipalities		
Rural	34.5	39.9
Semi-urbanized	35.9	39.4
Urbanized	33.9	35.3

Table 2. Distribution of marital status, smoking habits and overall frequency of meat consumption among men and women in the total cohort and in The Netherlands

	Men(%)		Women(%)	
Variable	Cohort	Netherlands	Cohort	Netherlands
Marital status				
Not married	3.8	6.2*	8.4	7.7*
Divorced	3.6	4.2	4.4	4.8
Married	88.9	85.3	69.9	68.1
Widowed	3.7	4.3	17.2	19.4
Smoking habits				
Never	9.3	4.0†	58.5	53.0†
Ex	48.8	51.0	20.7	27.0
Current	41.9	45.0	20.8	20.0
Meat consumption (freq. per week)				
0–1	1.9‡		3.6‡	
2–3	4.9		7.7	
4–5	24.4		29.1	
6–7	68.8		59.8	

*Age category 55–69 yr [54].
†Smoking habits in 1983, 51+ yr [55].
‡No large-scale reference data available in The Netherlands.

Table 3, indicating comparable intake estimates in the two studies.

These data indicate that the response to the baseline measurement has not adversely affected determinant distributions, in the light of etiological analyses. Although of less importance, it can also be concluded that no large deviations from representativeness with respect to these variables are evident.

FOLLOW-UP AND ANALYSIS ISSUES

As mentioned earlier, follow-up for cancer in this cohort of 120,852 subjects will consist of record linkage to PALGA and the cancer registries. As an alternative to a classical cohort analysis, the covariate histories of incident cases could also be compared to those of a control group in a nested case–control study [56, 57]. However, one would then need to wait until case occurrence for efficient matched sampling and subsequent standardized questionnaire processing for cases and control subjects. To overcome this problem, we employed a case–cohort (case–base) approach, as proposed by Miettinen [58] and Prentice [59], which offers the possibility of data processing during rather than after case ascertainment. In this approach, the denominator information of the rates (i.e. the accumulated person years of the entire cohort) is estimated using a subcohort of sufficient size, while cases are enumerated for the entire cohort (numerator information).

Required size of subcohort

Determination of the required subcohort size (3500) for testing the primary hypotheses in the case–cohort study was initially based on asymptotic relative efficiency comparisons for risk ratios. Efficiency results regarding rate ratios of Self and Prentice [60] had not yet been published at the time the decision on size had to be made. The asymptotic variances for the logarithm of the risk ratios estimated from the classical full cohort design (denoted by VCO) and from the case–cohort design (VCC) were calculated

Table 3. Mean caloric intake and its contributors among men and women in the cohort, as calculated from 9-day dietary records, and in The Netherlands (2-day dietary record)

	Men		Women	
Variable	Cohort ($n = 60$)	Netherlands* ($n = 431$)	Cohort ($n = 52$)	Netherlands* ($n = 460$)
Caloric intake (kcal)	2408	2564	1981	1946
Fat (% energy)	40.0	41.2	40.7	41.1
Protein (% energy)	14.1	13.7	14.4	14.9
Carbohydrates (% energy)	40.9	40.2	42.1	41.2
Alcohol (% energy)	5.0	5.0	2.8	2.7

*Ministry of Welfare, Public Health and Culture [30].

under simplifying assumptions: no competing risks, negligible loss to follow-up, and for a single dichotomous exposure variable. With S being the ratio of the subcohort size to the expected number of cases, VCO and VCC were calculated for a range of values of relative risk (RR; 0.1–10), control exposure probability (α; 2–90%), expected 5-year cumulative incidence (CI; 0.2–2%) and S (1–25). As an example of a typical situation for a dietary exposure, Fig. 3 shows a plot of VCC against values of S for RR = 2, α = 33%, CI = 0.2% (female rectum cancer), CI = 0.4% (male stomach) and CI = 2% (male lung), respectively. This figure illustrates that the variance estimate (or confidence interval) for the less common cancers will never be as small as that for lung cancer. The graph further indicates that for female rectum cancer the decrease in variance is minimal when S is increased over 16, while for male stomach and lung cancer this value of S is approximately 8 and 2, respectively.

For the various tumor sites, the relative efficiency VCO/VCC [27, 61] was then considered. Figure 4 shows VCO/VCC as a function of RR for α = 33%, CI = 0.4% and S = 1,2,4,8,16 and 25. Figure 4 indicates that S-values of 8 or higher are clearly sufficient over the entire range of RR-values. Similar results were obtained for the other tumor sites.

After considering the relative efficiencies under various conditions for various subcohort sizes and the added cost of processing additional questionnaires, we decided to choose a random subcohort of 3500 subjects. For most tumor sites, S-values or more are attained with this subcohort (e.g. 9 for male stomach and 16 for female rectum cancer); for male lung and female breast cancer the value of S would be 1.5 and 2, respectively. While the associated efficiencies for the latter tumor sites would be 50–60% for various combinations of α and RR, it should be kept in mind that the confidence intervals would still be much smaller than for less frequent cancer sites (see Fig. 3).

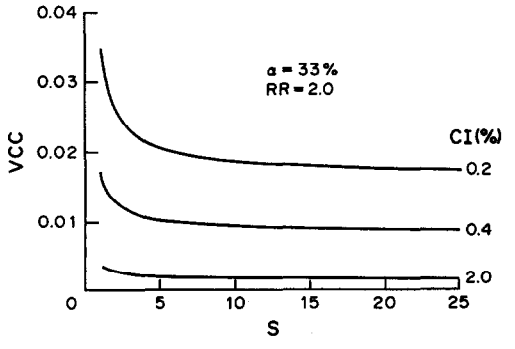

Fig. 3. The variance of logRR as a function of the ratio (subcohort:cases) for 3 cumulative incidence rates, using the case–cohort method. VCC, variance of logRR with the case–cohort method; S, ratio of subcohort:cases; CI, cumulative incidence; α, control exposure probability; RR, relative risk.

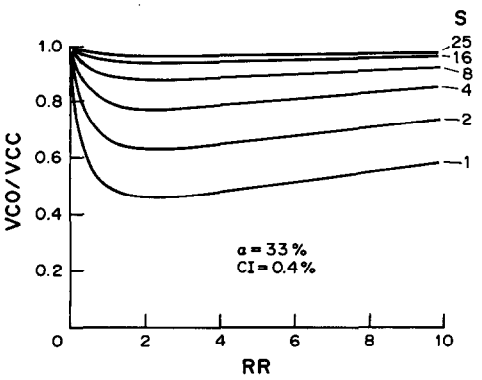

Fig. 4. The relative efficiency of the case–cohort vs full cohort analysis as a function of RR, for various subcohort sizes. VCO, variance of logRR with the full cohort method; VCC, variance of logRR with the case–cohort method; S, ratio of subcohort:cases; CI, cumulative incidence; α, control exposure probability; RR, relative risk.

To check the efficiency of the chosen subcohort size with regard to rate ratios and more realistic assumptions, parametric relative rate regression models for case–cohort studies were formulated. Based on these models, simulation studies were performed; the results were in accordance with those concerning the risk ratio.

DISCUSSION

We have started a prospective study on diet and cancer in a general population cohort of 120,852 men and women, in which determinant information from questionnaires and from toenail clippings is analyzed together with cancer incidence, using the case–cohort method. Conducting a study among the general population has the disadvantage of possible incomplete control for confounding of e.g. occupation as opposed to cohorts that are restricted in this sense (e.g. [7]). On the other hand, when these confounders are measured accurately, it also provides an opportunity to evaluate their effect modification (e.g. of occupation). The choice is usually determined, however, by the availability of specific population rosters and the possibilities for follow-up. The presence of both municipal population registries and

cancer registries in The Netherlands offered the opportunity for efficient recruitment and follow-up of the present cohort.

Another aspect that contributed to the efficiency was the increase in determinant contrast in the cohort by the intentional over-representation of vegetarians, albeit to a small extent. The somewhat disappointing experience in the recruitment of these individuals illustrates the inefficiency of obtaining large samples with accurate data through advertisements, as opposed to sampling from computerized population rosters with high quality data, needed when cancer follow-up is based on record linkage.

Loss to follow-up is the primary source of potential selection bias in prospective cohort studies (provided it is differential across determinant strata [62, 63]). Therefore it should be minimized, like in experimental studies. Hence, the study area and population in the present cohort study were chosen in a way to ensure sufficient follow-up coverage. Recruitment of a large general population cohort in the way described, implies an incomplete response to the baseline measurement. Bias in determinant distributions due to nonresponse has no serious implications for ratio estimates, even though respondents generally show lower mortality or disease experience during follow-up than non-respondents (e.g. [64, 65]). In studies that have addressed the issue of nonresponse, odds ratio estimates were not significantly different between participants and non-participants, although both groups exhibited (largely independent) differences in determinant distributions and disease experience [65, 66]. In fact, the distribution of risk factors may even become more favorable for etiologic studies due to response at baseline. This potentially increased efficiency is also why intentional overrepresentation of vegetarians was pursued in the present study, and why in an experimental situation subjects are allocated equally to determinant strata. Data on demographic variables, smoking and dietary habits were presented indicating that the response did not adversely affect determinant distributions in the present cohort. To evaluate whether differential loss to follow-up occurs, we will compare the determinant profile of those lost to follow-up with other participants.

An elderly cohort was selected because dietary habits (and their contrasts) are stabilized, and such a cohort will yield sufficient cases for meaningful analyses within a reasonable time period. It can be argued that evaluation of nutritional determinants of cancer acting early in life [67] cannot be evaluated with this approach. Since it has been suggested that various dietary factors act in later stages of carcinogenesis and a large-scale study among a cohort of e.g. adolescents would be time-consuming with the need to consider intermediate (dietary) events also, this potential drawback was accepted. Together with other ongoing studies and studies that will investigate the role of diet in the earlier stages of carcinogenesis, this study will contribute to a better understanding of the type, timing and weight of the influence diet may have on human cancer development.

Acknowledgements—This work was supported by grants from The Netherlands Cancer Foundation and the Ministry of Welfare, Public Health and Cultural Affairs. The authors want to thank M. Bethlehem, H. Brants, C. van Deursen, G. Franssen, H. Hofhuis, A. Huldij, M. de Leeuw, H. van Montfort, A. Schiphorst, R. Schmeitz, H. Smit and G. Snijders for their assistance in the conduct of this study, and F. J. Kok for useful comments on this manuscript.

REFERENCES

1. Wynder EL, Gori GB. Contribution of the environment to cancer incidence: an epidemiologic exercise. **J Natl Cancer Inst** 1977; 58: 825–832.
2. Higginson J, Muir CS. Environmental carcinogenesis: misconceptions and limitations to cancer control. **J Natl Cancer Inst** 1979; 63: 1291–1298.
3. Doll R, Peto R. The causes of cancer. **J Natl Cancer Inst** 1981; 66: 1191–1308.
4. Zaridze DG, Muir CS, McMichael AJ. Diet and cancer: value of different types of epidemiological studies. **Nutr Cancer** 1985; 7: 155–166.
5. Graham S. Results of case–control studies on diet and cancer in Buffalo, New York. **Cancer Res** 1983; 43 (Suppl.): 2409s–2413s.
6. Kvale G, Bjelke E, Gart JJ. Dietary habits and lung cancer risk. **Int J Cancer** 1983; 31: 397–405.
7. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE. Dietary fat and the risk of breast cancer. **N Engl J Med** 1987; 316: 22–28.
8. Phillips RL, Snowdon DA. Association of meat and coffee use with cancers of the large bowel, breast and prostate among Seven-Day Adventists: preliminary results. **Cancer Res** 1983; 43 (Suppl.): 2403s–2408s.
9. Enstrom JE. Cancer mortality among Mormons in California during 1968–75. **J Natl Cancer Inst** 1980; 65: 1073–1082.
10. Kolonel LN, Nomura AMY, Hinds MW, Hirohata T, Hankin JH, Lee J. Role of diet in cancer incidence in Hawaii. **Cancer Res** 1983; 43(Suppl.): 2397s–2402s.
11. Hirayama T. Diet and cancer. **Nutr Cancer** 1979; 1: 67–81.
12. Miller AB, Howe GR, Wall C. The national study of breast cancer screening. **Clin Invest Med** 1982; 4: 227–258.

13. Van Noord PAH, Collette HJA, Maas MJ, De Waard F. Selenium levels in nails of premenopausal breast cancer patients assessed prediagnostically in a cohort-nested case-referent study among women screened in the DOM project. **Int J Epidemiol** 1987; 16 (Suppl.): 318–322.
14. Salonen JT, Salonen R, Lappetelainen R, Maenpaa PH, Alfthan G, Puska P. Risk of cancer in relation to serum concentrations of selenium and vitamins A and E: matched case–control analysis of prospective data. **Br Med J** 1985; 290: 417–420.
15. Kok FJ, De Bruin AM, Hofman A, Vermeeren R, Valkenburg HA. Is serum selenium a risk factor for cancer in men only? **Am J Epidemiol** 1987; 125: 12–16.
16. Stellman SD, Garfinkel L. Artificial sweetener use and one-year weight change among women. **Prev Med** 1986; 15: 195–202.
17. Paganini-Hill A, Chao A, Ross RK, Henderson BE. Vitamin A, beta-carotene, and the risk of cancer: a prospective study. **J Natl Cancer Inst** 1987; 79: 443–448.
18. Jones DY, Schatzkin A, Green SB, Block G, Brinton LA, Ziegler RG, Hoover R, Taylor PR. Dietary fat and breast cancer in the National Health and Nutrition Examination Survey—I. Epidemiologic Follow-up Study. **J Natl Cancer Inst** 1987; 79: 465–471.
19. Nomura A, Heilbrun LK, Stemmermann GN. Prospective study of coffee consumption and the risk of cancer. **J Natl Cancer Inst** 1986; 76: 587–590.
20. Kromhout D, Bosschieter EB, De Lezenne Coulander C. Dietary fibre and 10-year mortality from coronary heart disease, cancer, and all causes; the Zutphen Study. **Lancet** 1982; ii: 518–521.
21. Van den Brandt PA, Goldbohm RA, van 't Veer P, Hermus RJJ, Sturmans F. Dietary habits and the aetiology of cancer. **Int J Epidemiol** 1988; 17: 472.
22. Palmer S. Diet, nutrition and cancer. **Prog Food Nutr Sci** 1985; 9: 283–341.
23. Willett WC, MacMahon B. Diet and cancer–an overview (two parts). **N Engl J Med** 1984; 310: 633–639 and 697–703.
24. Muir C, Waterhouse J, Mack T, Powell J, Whelan S, Eds. **Cancer Incidence in Five Continents,** Vol. V. Lyon: IARC Sci. Publ.; 1987: 88.
25. Central Bureau of Statistics. Kankermorbiditeit en mortaliteit 1984–1985. **Maandbericht Gezondheid** 1987; 6: 5–21.
26. Central Bureau of Statistics. Sterfte, 1982–1986. **Maandstatistiek Bevolking** 1987; 9: 36–39.
27. Breslow NE, Day NE. **Statistical Methods in Cancer Research. II. The Design and Analysis of Cohort Studies.** Lyon: IARC Sci. Publ.; 1987: 82.
28. Cox DR. Regression models and life tables (with discussion). **J R Stat Soc B** 1972; 34: 187–220.
29. National Academy of Sciences, National Research Council, Committee on Diet, Nutrition, and Cancer. **Diet, Nutrition, and Cancer.** Washington: National Academy Press; 1982.
30. Ministry of Welfare, Public Health and Culture, Ministry of Agriculture and Fisheries. **Wat eet Nederland?** Rijswijk, The Netherlands, 1988.
31. Van Beresteyn ECH, Van't Hof MA, De Waard H, Dekker PR, Neeter R, Winkeldermaat HJ, Visser RM, Schaafsma G, Van Schiak M, Duursma SA. Design and data quality of a mixed longitudinal study to elucidate the role of dietary calcium and phosphorus in bone mineralization in pre-, peri- and post-menopausal women. **Am J Clin Nutr** 1986; 4: 538–548.
32. Post GB. Nutrition in adolescence. Dissertation, Wageningen, 1989.
33. Central Bureau of Statistics. Kanker 1978–1979; morbiditeit en mortaliteit. **Maandstatistiek Bevolking Volksgezondheid** 1981; 29: 46–61.
34. Haenszel W, Kurihara M, Locke FB, Shimuzu K, Segi M. Stomach cancer in Japan. **J Natl Cancer Inst** 1976; 56: 265–274.
35. Higginson J. Etiological factors in gastrointestinal cancer in man. **J Natl Cancer Inst** 1966; 37: 527–545.
36. Hebert JR, Miller DR. Methodologic considerations for investigating the diet-cancer link. **Am J Clin Nutr** 1988; 47: 1068–1077.
37. Liu K, Stamler J, Dyer A, McKeever P, McKeever J. Statistical methods to assess and minimize the role of intraindividual variability in obscuring the relationship between dietary lipids and serum cholesterol. **J Chron Dis** 1978; 31: 399–418.
38. Oiso T. Incidence of stomach cancer and its relation to dietary habits and nutrition in Japan between 1900 and 1975. **Cancer Res** 1975; 35: 3254–3258.
39. Netherlands Nutrition Council. Richtlijnen goede voeding. **Voeding** 1986; 47: 159–181.
40. Rohan TE, Potter JD. Retrospective assessment of dietary intake. **Am J Epidemiol** 1984; 120: 876–887.
41. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. **Am J Epidemiol** 1985; 122: 51–65.
42. Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. **JADA** 1987; 87: 43–47.
43. Jain MG, Harrison L, Howe GR, Miller AB. Evaluation of a self-administered dietary questionnaire for use in a cohort study. **Am J Clin Nutr** 1982; 36: 931–935.
44. Clayton DG. using test-retest reliability data to improve relative risk estimates: an application of latent class analysis. **Stat Med** 1985; 4: 445–455.
45. Marshall JR, Graham S. Use of dual responses to increase validity of case–control studies. **J Chron Dis** 1984; 37: 125–136.
46. Bausch-Goldbohm RA, van den Brandt PA, van 't Veer P, Sturmans F, Hermus RJJ. Results of the methodological study for the design of a simplified, self-administered questionnaire. In: Riboli E, Saracci R, Eds. **Diet, Hormones and Cancer: Methodological Issues for Prospective Studies.** Lyon: IARC Technical Report; 1988: 4: 79–89.
47. Hodges RE, Hood J, Canham JE, Sauberlich HE, Baker EM. Clinical manifestations of ascorbic acid deficiency in man. **Am J Clin Nutr** 1971; 24: 432–443.
48. Robinson MF, Godfrey PJ, Thomson CD, Rea HM, Van Rij AM. Blood selenium and glutathione peroxidase activity in normal subjects and in surgical patients with and without cancer in New Zealand. **Am J Clin Nutr** 1979; 32: 1477–1485.
49. Wald N, Boreham J, Bailey A. Serum retinol and subsequent risk of cancer. **Br J Cancer** 1986; 54: 957–961.
50. Wald NJ, Thompson SG, Densem JW, Boreham J, Bailey A. Serum vitamin E and subsequent risk of cancer. **Br J Cancer** 1987; 56: 69–72.
51. Morris JS, Stampfer MJ, Willett W. Dietary selenium in humans. Toenails as an indicator. **Biol Trace Elem Res** 1983; 5: 529–537.
52. Schrauzer GN, White DA, Schneider CJ. Cancer mortality correlation studies IV: Associations with dietary intakes and bloodlevels of certain trace elements, notably Se-antagonists. **Bioinorg Chem** 1977; 7: 35–54.
53. Van den Brandt PA, Bausch-Goldbohm RA, van 't Veer P, Hermus RJJ, Sturmans F. The Dutch prospective cohort study on diet and cancer. In: Riboli E, Saracci R, Eds. **Diet, Hormones and Cancer: Methodological Issues for Prospective Studies.** Lyon: IARC Technical Report; 1988: 4: 18–27.

54. Central Bureau of Statistics. **Statistisch Zakboek** 1986. The Hague: Staatsuitgeverij; 1986.
55. Van Reek J, MJ Drop, H Adriaanse. Stoppen met roken sinds 1958. **Tydschrift Alcohol Drugs** 1985; 11: 168–173.
56. Mantel N. Synthetic retrospective studies and related topics. **Biometrics** 1973; 29: 479–486.
57. Thomas DC. Addendum to a paper by Liddell FDK, McDonald JC & Thomas DC. **J R Stat Soc A** 1977; 140: 483–485.
58. Miettinen OS. **Theoretical Epidemiology. Principles of Occurrence Research in Medicine.** New York: Wiley; 1985.
59. Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. **Biometrika** 1986; 73: 1–12.
60. Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case–cohort studies. **Ann Stat** 1988; 16: 64–81.
61. Whittemore AS, McMillan A. Analysing occupational cohort data: application to US uranium miners. In: Prentice RL, Whittemore AS, Eds. **Environmental Epidemiology: Risk Assessment.** Philadelphia: SIAM; 1982: 65–81.
62. Greenland S. Response and follow-up bias in cohort studies. **Am J Epidemiol** 1977; 106: 184–187.
63. Kleinbaum DG, Kupper LL, Morgenstern H. **Epidemiologic Research: Principles and Quantitative Methods.** Belmont, California: Lifetime Learning; 1982.
64. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. **Br Med J** 1964; 1: 1399–1410.
65. Heilbrun LK, Nomura A, Stemmermann GN. The effects of nonresponse in a prospective study of cancer. **Am J Epidemiol** 1982; 116: 353–363.
66. Austin MA, Criqui MH, Barrett-Connor E, Holdbrook MJ. The effect of response bias on the odds ratio. **Am J Epidemiol** 1981; 114: 137–143.
67. Ross MH, Lustbader ED, Bras G. Dietary practices of early life and spontaneous tumors of the rat. **Nutr Cancer** 1982; 3: 150–167.